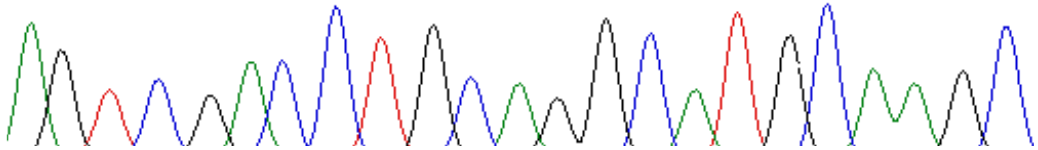


Genome Technologies

Dr Mick Jones

MRC CSC Genomics Laboratory

Hammersmith Hospital Campus

Imperial College London

MRC Clinical Sciences Centre

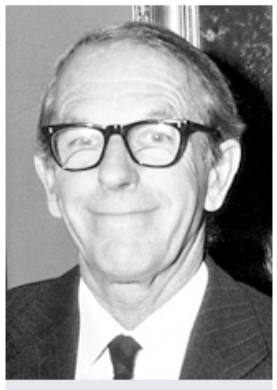
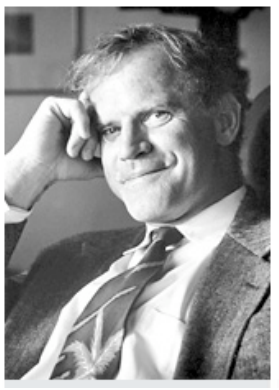










J. Mol. Biol. (1971) **56**, 341-361

Studies on Polynucleotides

XCVI. Repair Replication of Short Synthetic DNA's as catalyzed by DNA Polymerases

K. Kleppe, E. Ohtsuka, R. Kleppe, I. Molineux
and H.G. Khorana

Institute for Enzyme Research of the University of Wisconsin
Madison, Wisc. 53706, U.S.A.
(Received 20 July 1970)

The principles for extensive synthesis of the duplexed tRNA genes which emerge from the present work are the following. The DNA duplex would be denatured to form single strands. This denaturation step would be carried out in the presence of a sufficiently large excess of the two appropriate primers. Upon cooling, one would hope to obtain two structures, each containing the full length of the template strand appropriately complexed with the primer. DNA polymerase will be added to complete the process of repair replication. Two molecules of the original duplex should result. The whole cycle could be repeated, there being added every time a fresh dose of the enzyme. It is however, possible that upon cooling after denaturation of the DNA duplex, renaturation to form the original duplex would predominate over the template-primer complex formation. If this tendency could not be circumvented by adjusting the concentrations of the primers, clearly one would have to resort to the separation of the strands and then carry out repair replication. After every cycle of repair replication, the process of strand separation would have to be repeated. Experiments based on these lines of thought are in progress.

April 25, 1953 NATURE

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

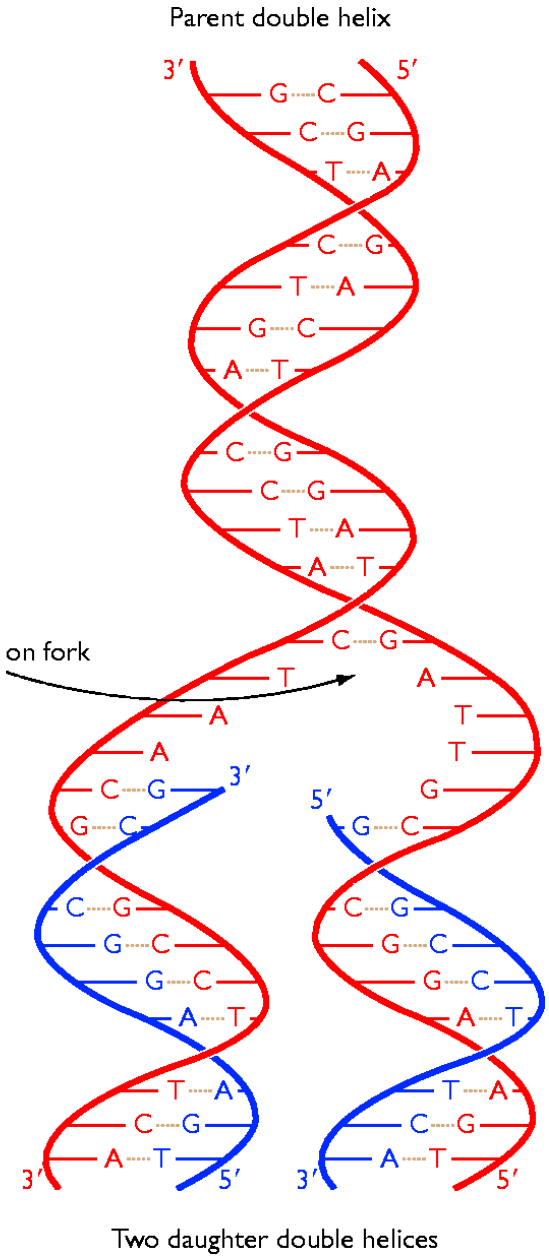
A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

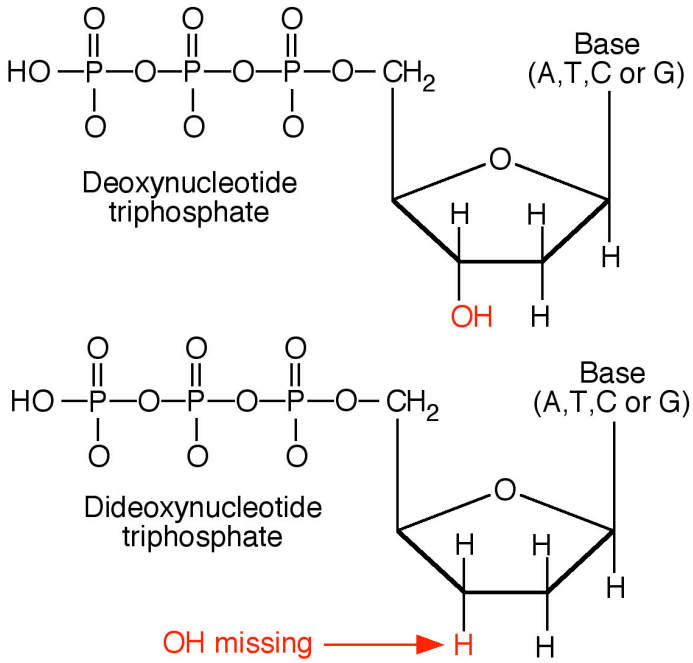
J. D. WATSON
F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.



Chain Termination

Fred Sanger



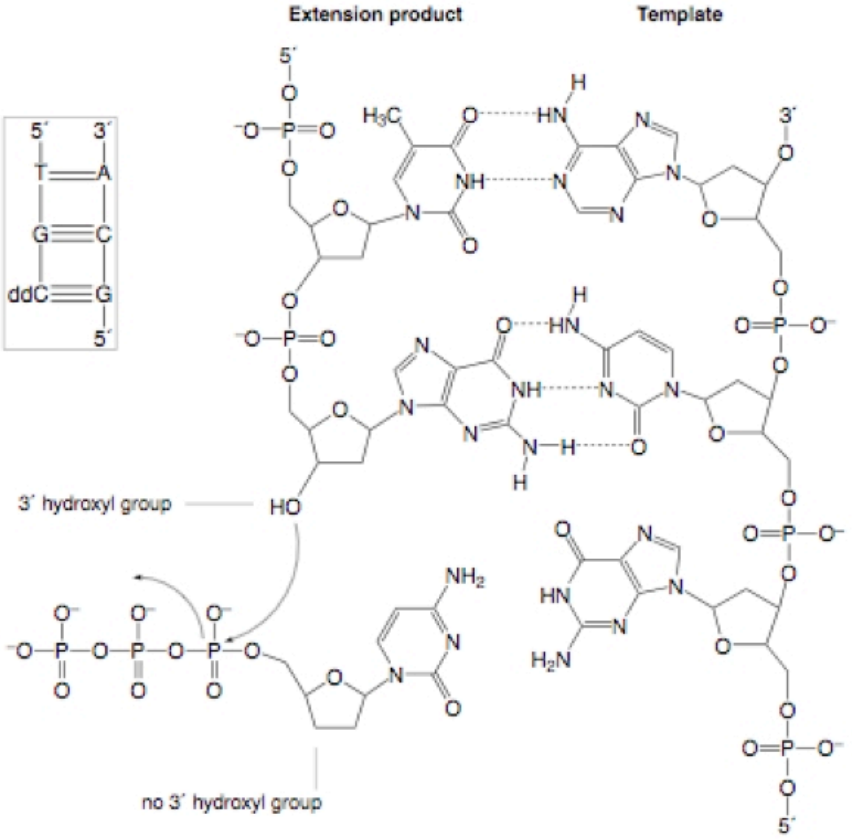
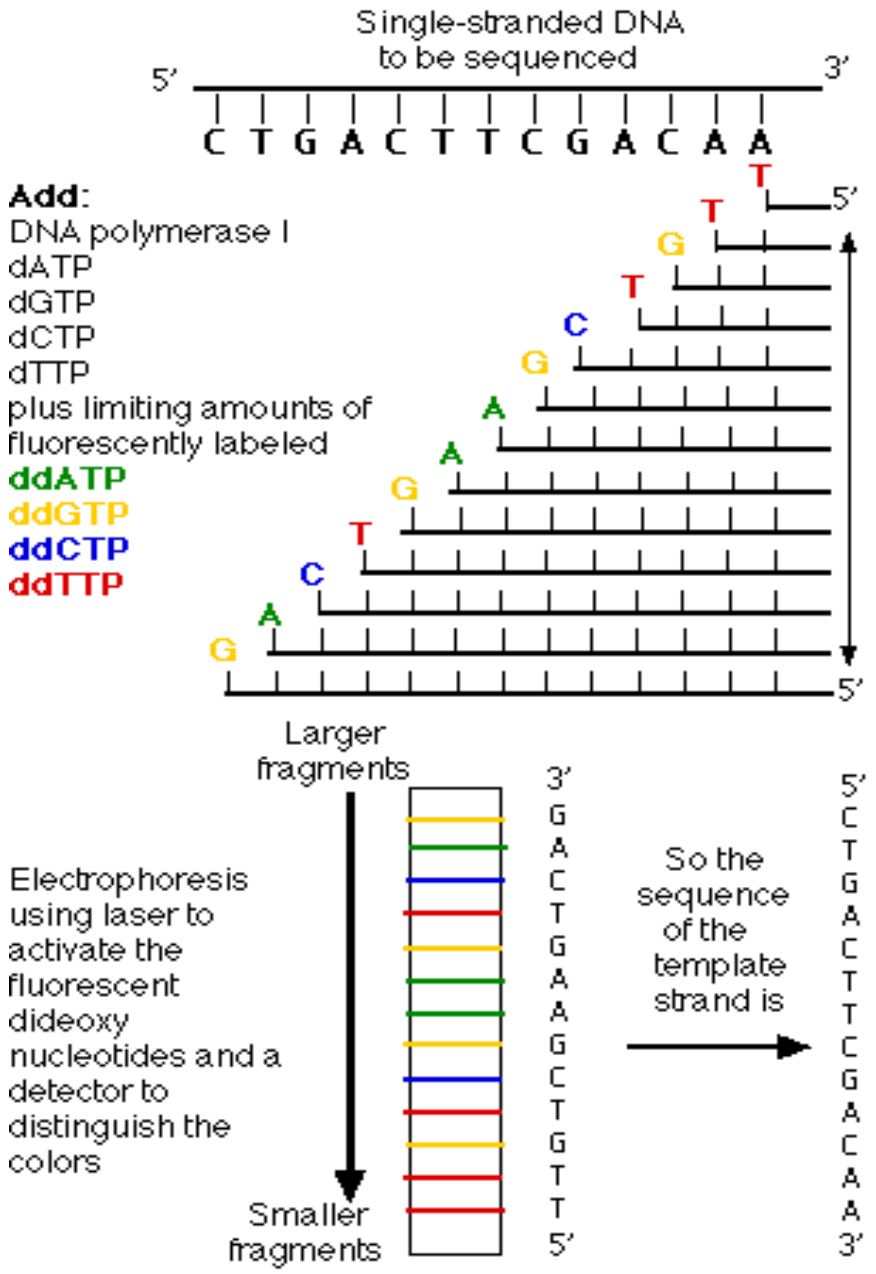


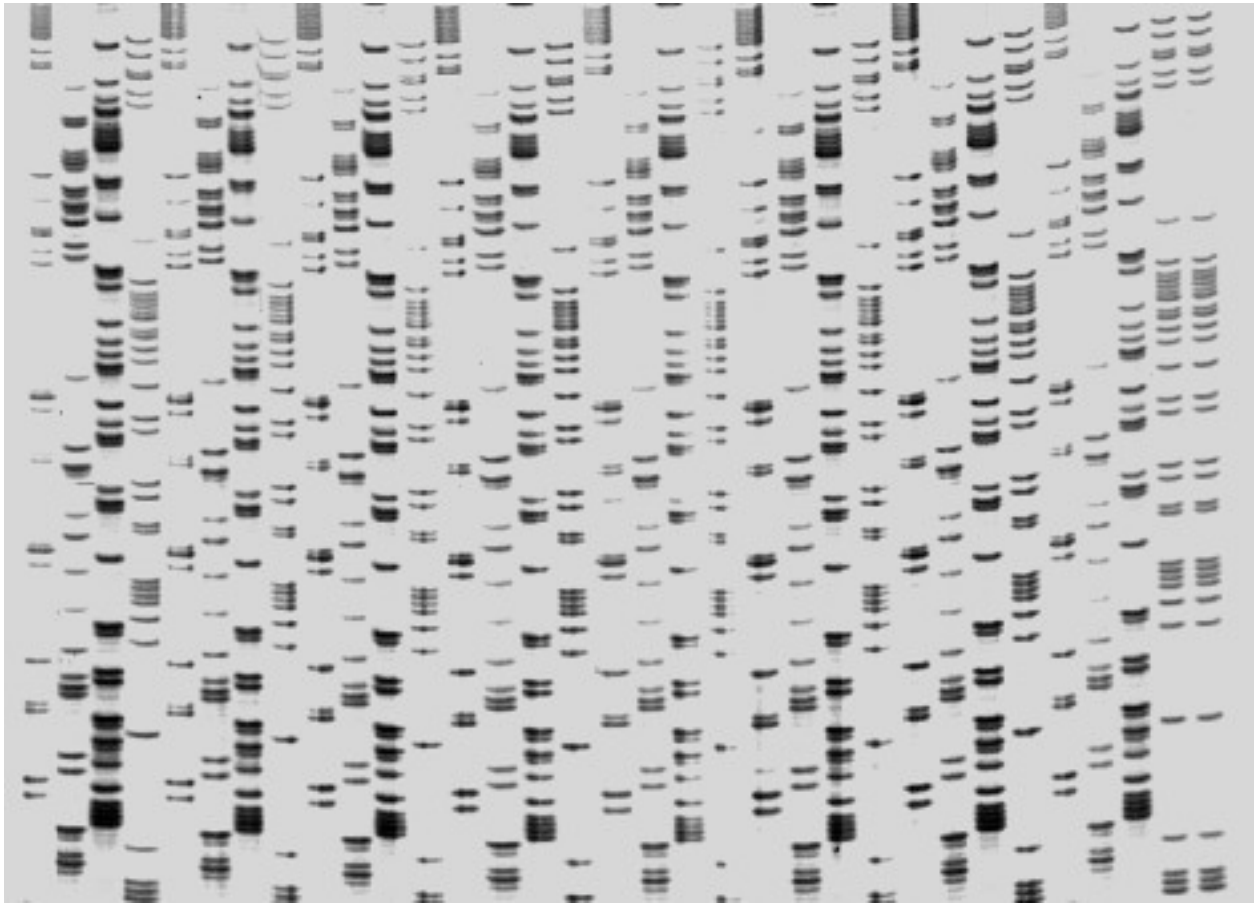
Figure 1-1 DNA strand synthesis by formation of phosphodiester bonds. The chain is terminated by the use of dideoxycytidine triphosphate (ddC) in place of deoxycytidine triphosphate (dCTP). The inset shows a schematic representation of the process.

DNA Sequencing Methodology

Sanger Sequencing or Dideoxy Chain Terminator Sequencing



Radioactive Sequencing – 4 Lane Technology



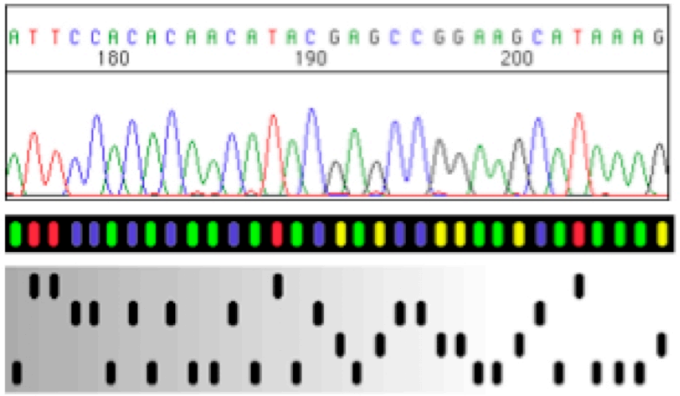
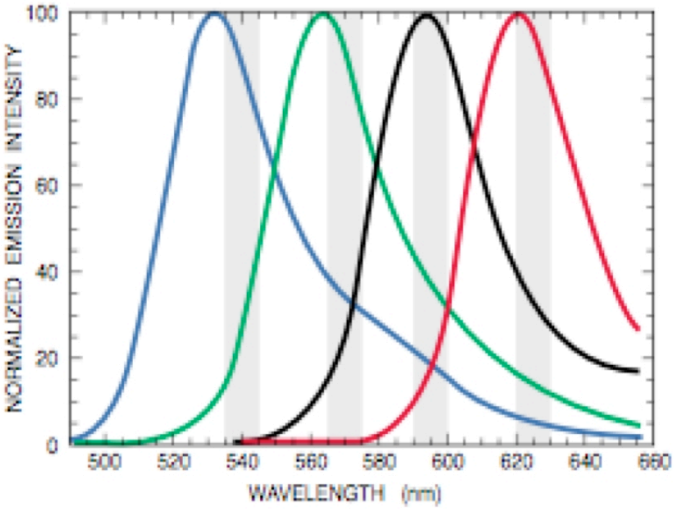


Figure 1-2 Four-color/one-lane fluorescent sequencing vs. one-color/four-lane method such as radioactive sequencing



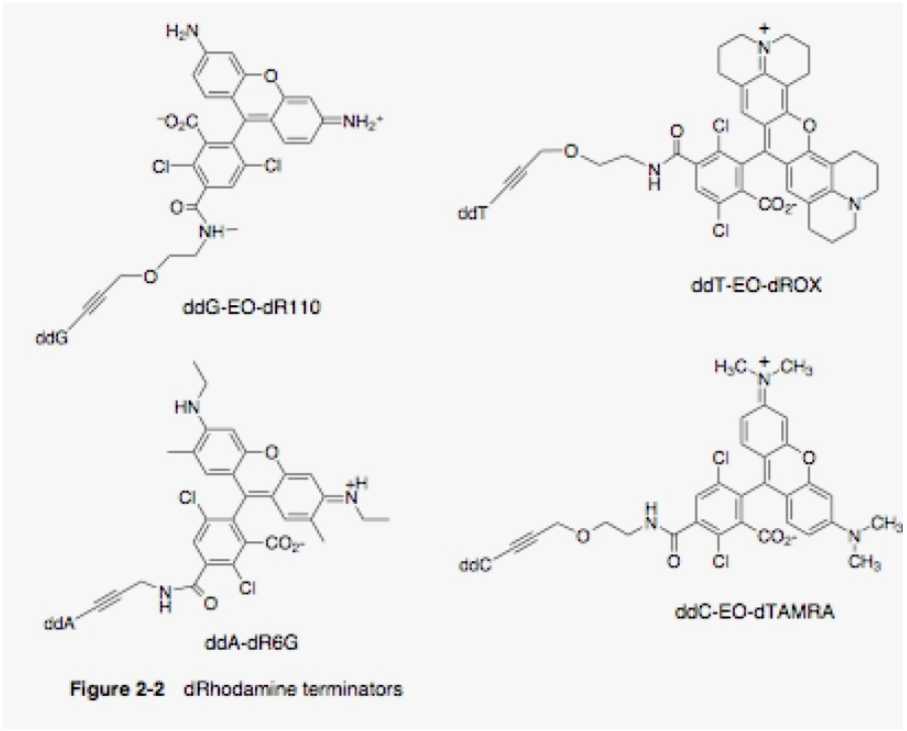
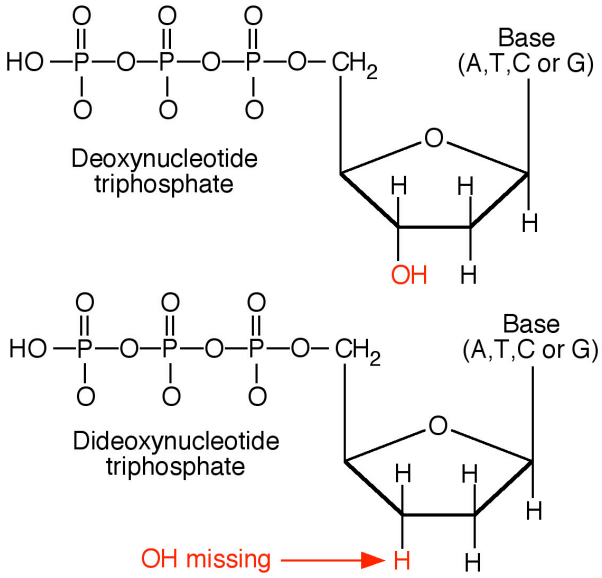


Figure 2-2 dRhodamine terminators



DNA Read Length Limitation

DNA Sequencing works because of denaturing gel matrix

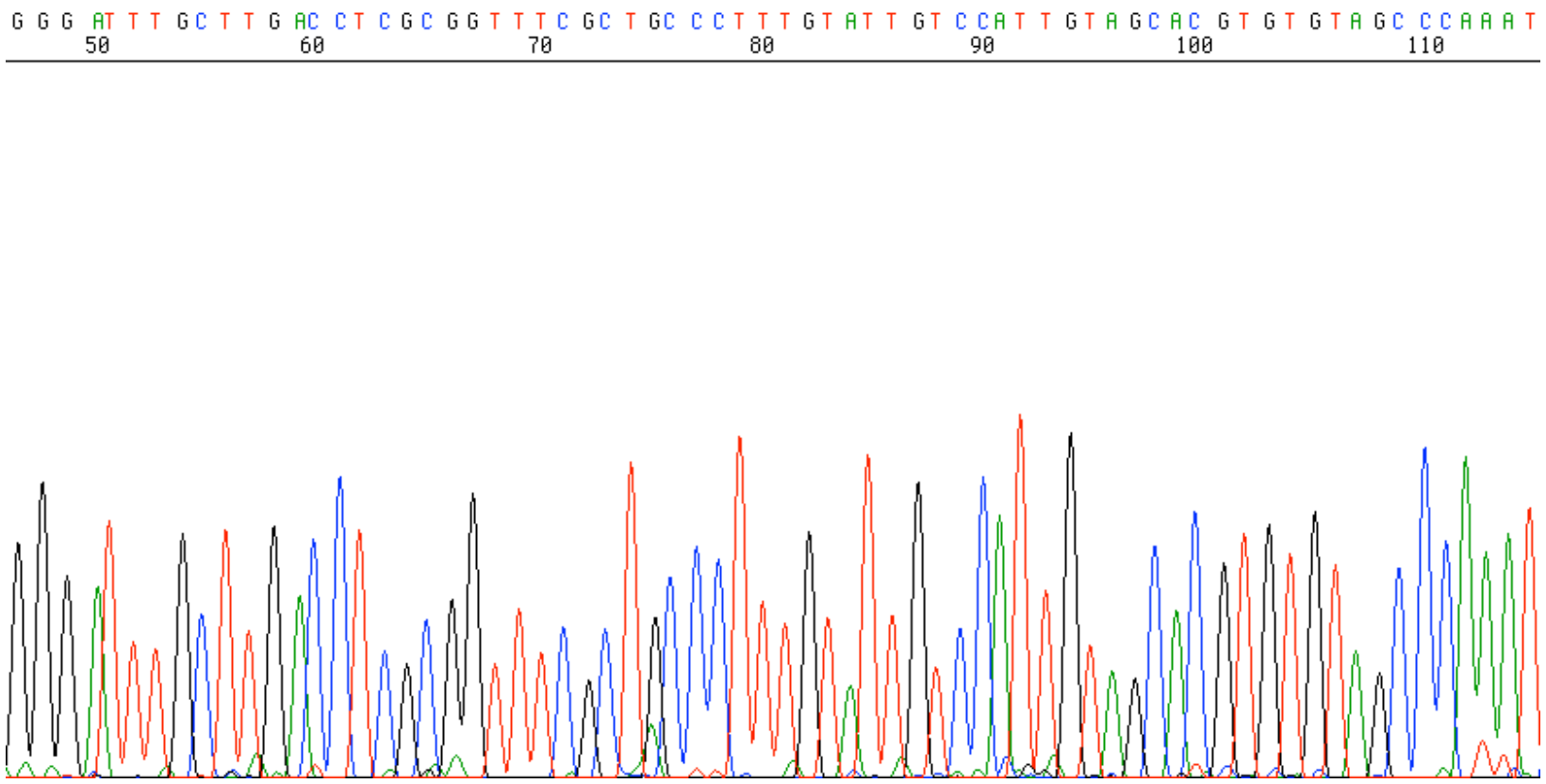
Radioactive method - 8 M Urea polyacrylamide gels

Fluorescent methods - capillary electrophoresis

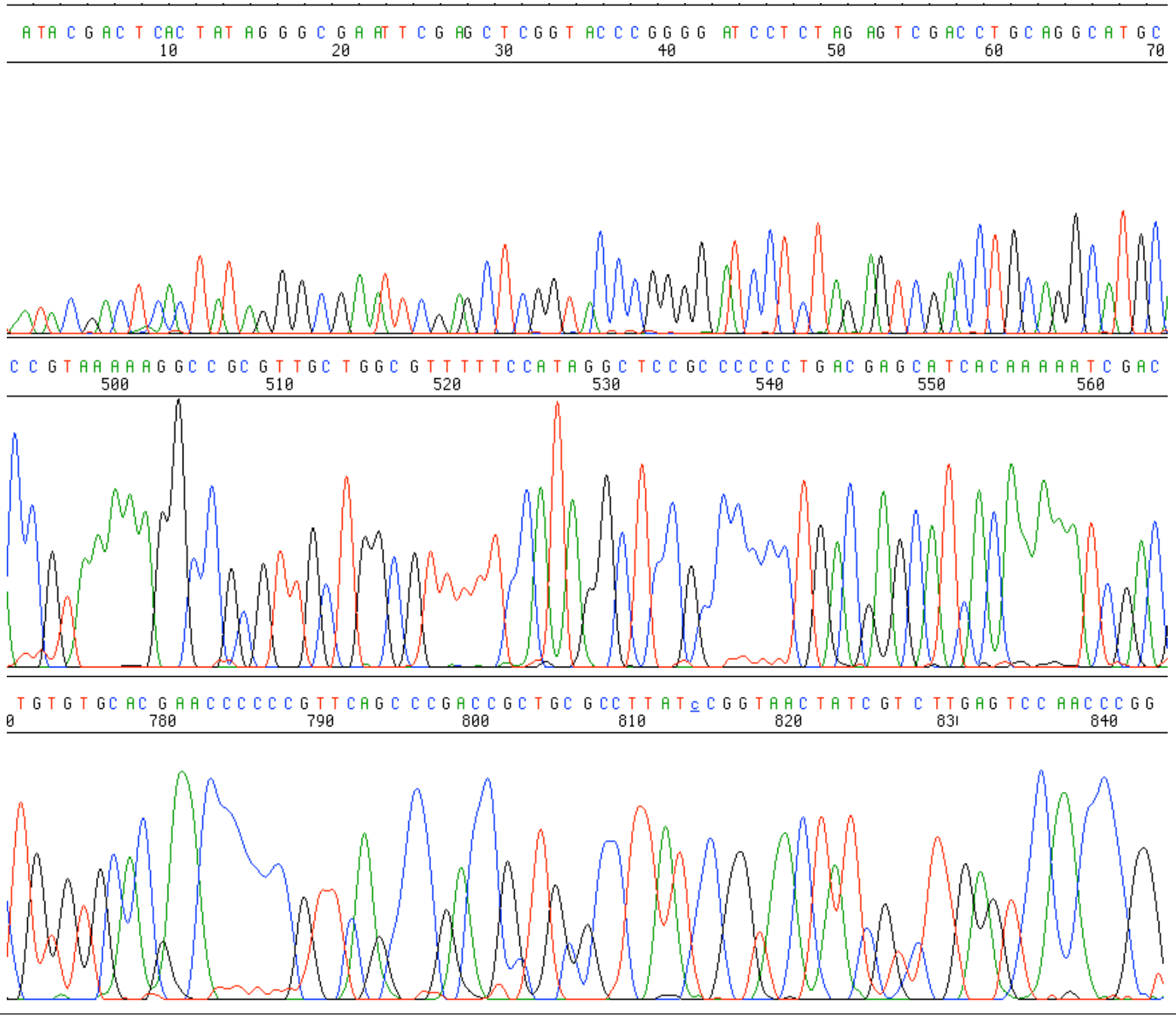
Allows separation of DNA fragments differing by 1 base in size
From 1 to 1,000 bases

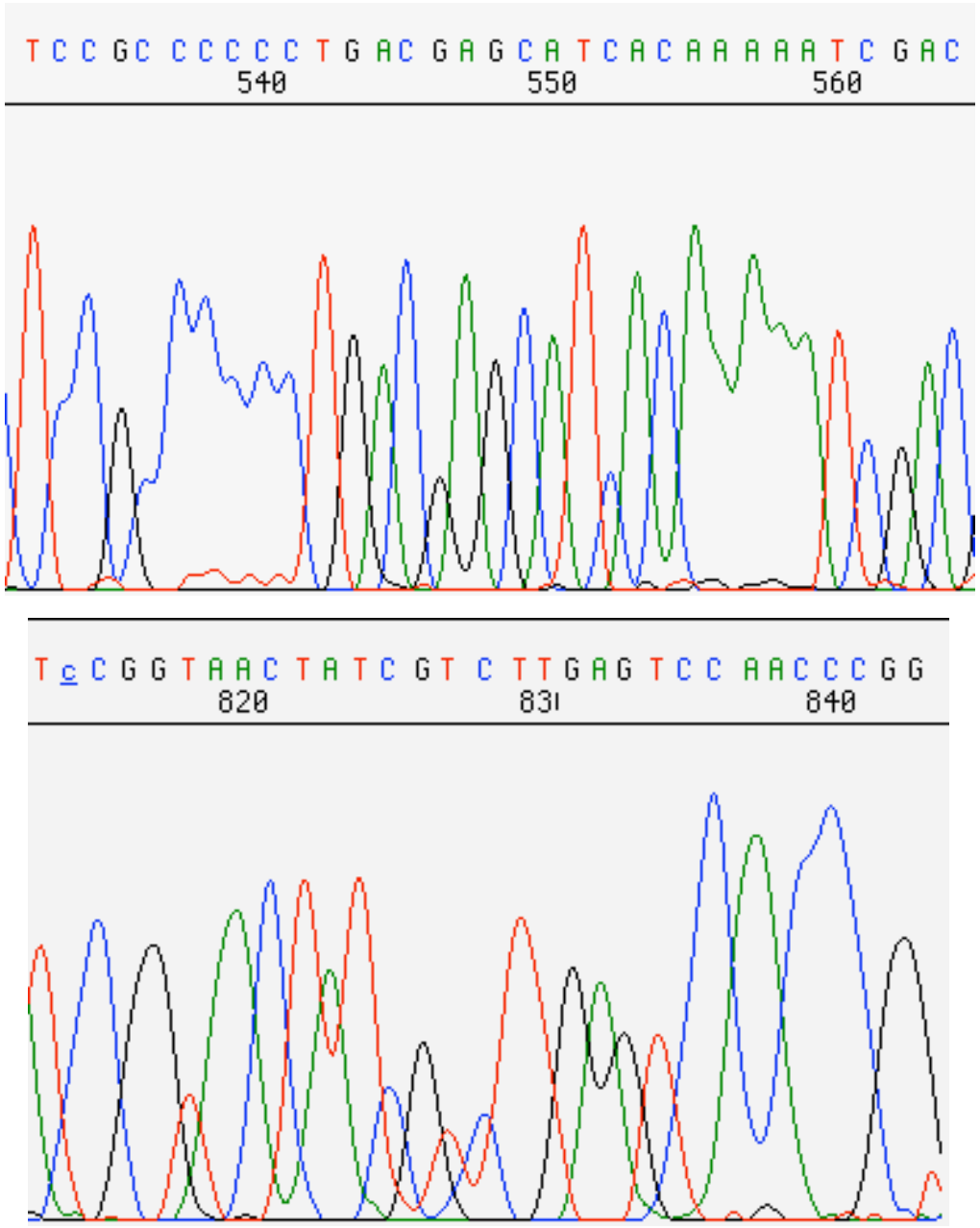
The larger the fragment, the less the separation.

DNA Sequence ‘Trace’ Chromatogram



Smaller fragments, good base separation


ATA C G AC T CAC T AT AG G G C G A AT T C G AG C T C GG T AC C C GG G G AT C C T C T AG AG T C G AC C T G C AG G C A T G C
10
20
30
40
50
60
70
C C G T AA A A A G G C C G C G T T G C T G G C G T T T T T C C A T A G G C T C C G C C C C C T G AC G A G C A T C A C A A A A A T C G A C
500
510
520
530
540
550
560
T G T G T G C AC G A A C C C C C C C G T T C A G C C C G A C C G C T G C G C C T T A T c C G G T A A C T A T C G T C T T G A G T C C A A C C C G G
0
780
790
800
810
820
830
840


T C C GC C C C C C T G AC G A G C A T C A C A A A A A T C G A C
540
550
560
T c C G G T A A C T A T C G T C T T G A G T C C A A C C C G G
820
831
840

ABI PRISM® 3700*xl* DNA Analyzer

Bench top instrument

Analyse 96 samples in 1 hours
~800 bases/sample

Thus in 24 hours can generate
~1,000,000 bases of sequence

Cost: £3.60/sample

New DNA Sequencing Technologies

Current Sanger methods valuable for small scale projects

But

Expensive and time consuming for re-sequencing projects, i.e. sequencing another human genome

Sequencing human genome (3×10^9 bp) to 10 fold coverage required ~50,000,000 sequence reactions

One 3730*xl* DNA sequencer would take 25,000 days working 24 hours a day (*68.5 years*!)

New DNA Sequencing Technologies

New approaches to Sequencing large regions ~10^6 bp

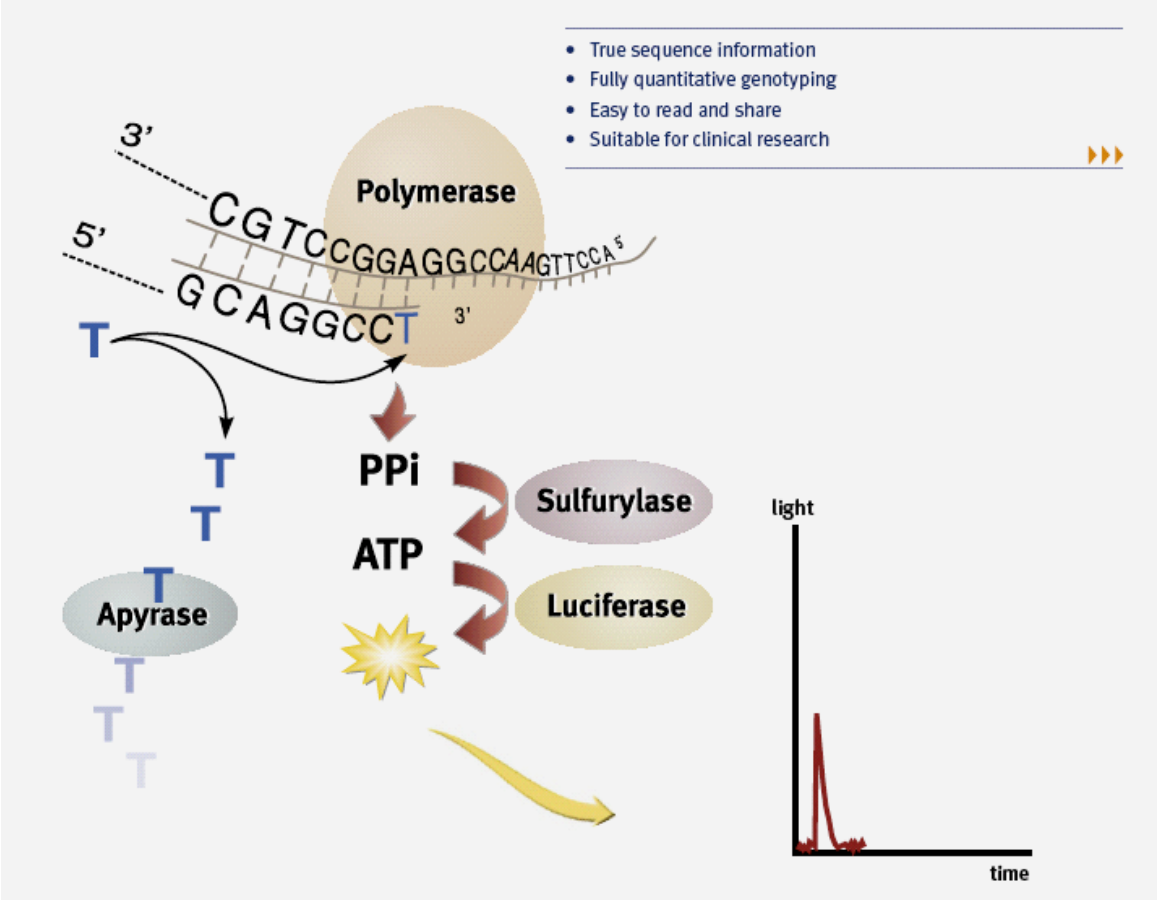
SBS - Sequence by Synthesis

AIM: A Human Genome Sequence for $1,000

A Human Genome Sequence in One Day

SBS – Sequence by Synthesis

Pyrosequencing - Synthesis one base at a time, with real time detection.

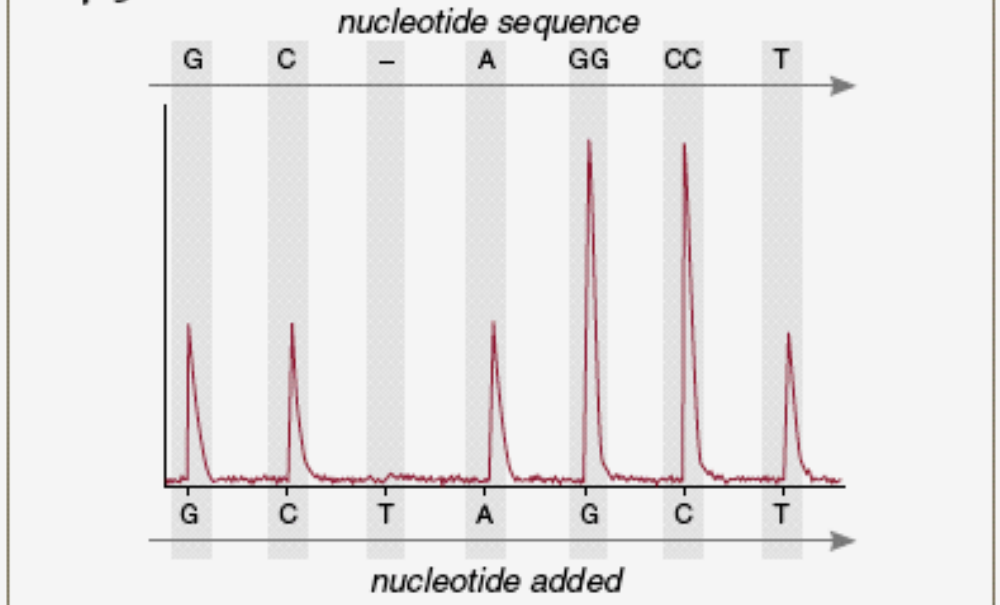


Step 4

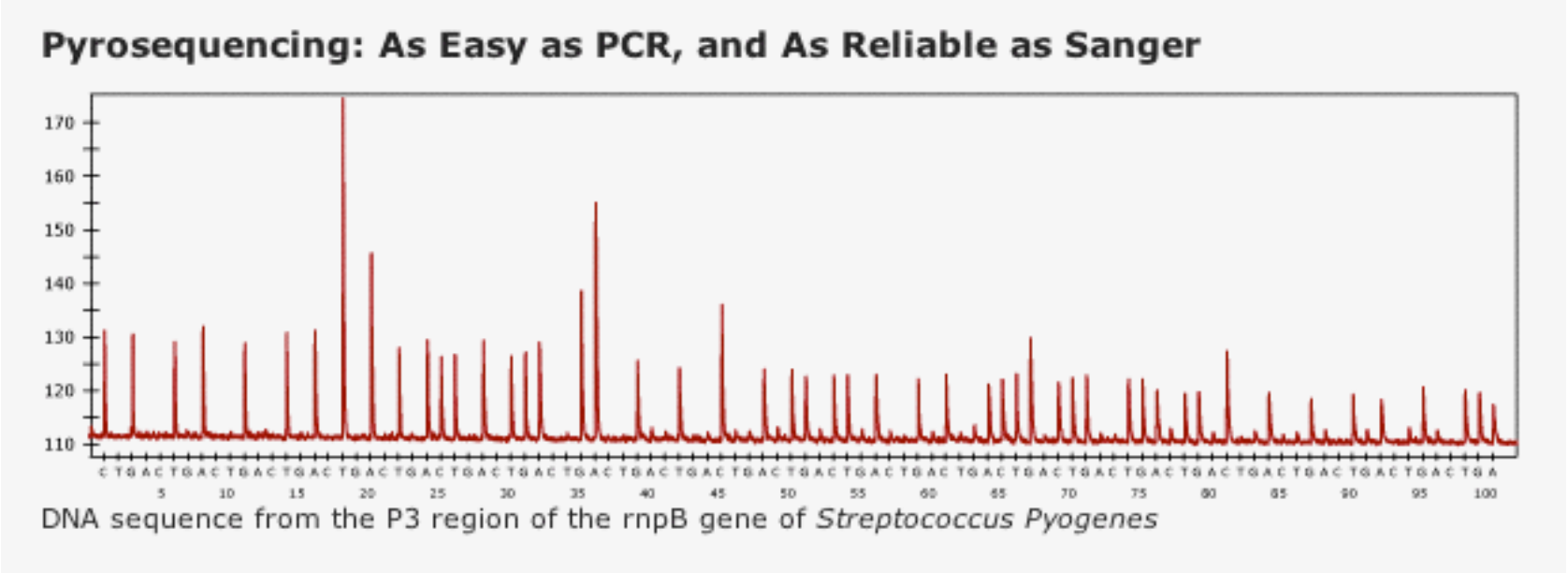
dNTP —(Apyrase)→ dNDP + dNMP + phosphate

ATP —(Apyrase)→ ADP + AMP + phosphate

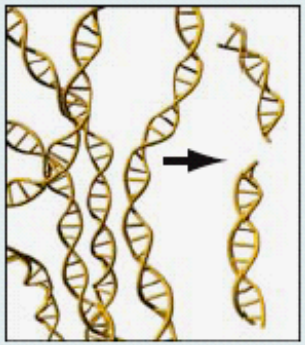
Step 5

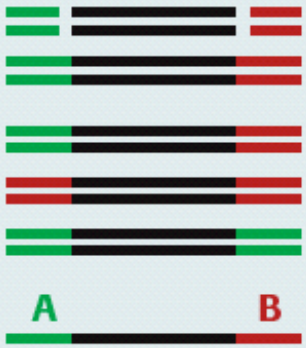


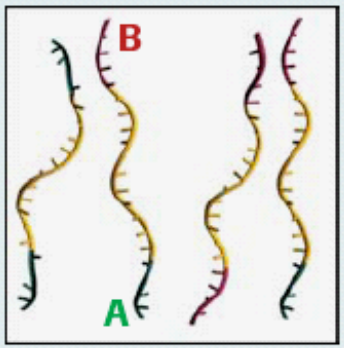
Read ~400 bases per sequence run



But this is smaller than Sanger Shotgun >700 bases of sequence

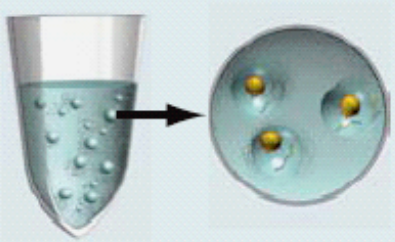




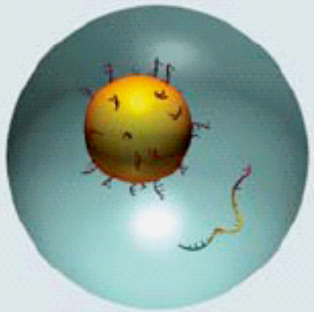


- Genome fragmented by nebulization
- No cloning; no colony picking
- sstDNA library created with adaptors
- A/B fragments selected using avidin-biotin purification

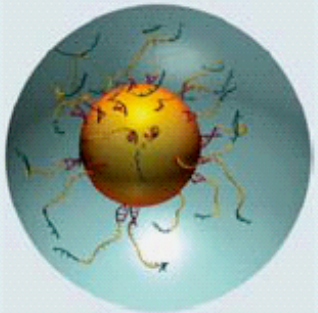
gDNA → **sstDNA library**



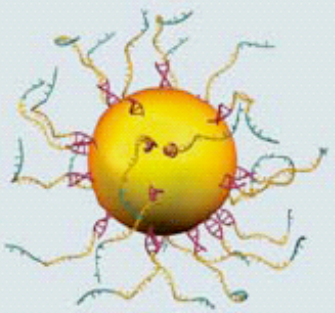
Anneal sstDNA to an excess of DNA Capture Beads



Emulsify beads and PCR reagents in water-in-oil microreactors



Clonal amplification occurs inside microreactors



Break microreactors, enrich for DNA-positive beads

sstDNA library → **Clonally-amplified sstDNA attached to bead**

Genome Sequencer FLX

Titanium Upgrade

- >1,000,000 reads/run
- >400 base read length/bead
- >400 million bases/run (10 hours)

The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler[1]*, Maithreyan Srinivasan[2]*, Michael Egholm[2]*, Yufeng Shen[1]*, Lei Chen[1], Amy McGuire[3], Wen He[2], Yi-Ju Chen[2], Vinod Makhijani[2], G. Thomas Roth[2], Xavier Gomes[2], Karrie Tartaro[2]†, Faheem Niazi[2], Cynthia L. Turcotte[2], Gerard P. Irzyk[2], James R. Lupski[4,5,6], Craig Chinault[4], Xing-zhi Song[1], Yue Liu[1], Ye Yuan[1], Lynne Nazareth[1], Xiang Qin[1], Donna M. Muzny[1], Marcel Margulies[2], George M. Weinstock[1,4], Richard A. Gibbs[1,4] & Jonathan M. Rothberg[2]†

Table 3 | SNPs matching HGMD mutations causing disease or other phenotypes

HGMD accession	Chromosome	Coordinate	HUGO symbol	Gene name	Cytogenetic	Phenotype	Zygosity
CM003589	1	97937679	DPYD	Dihydropyrimidine dehydrogenase	1q22	Dihydropyrimidine dehydrogenase deficiency	Heterozygous
CM950484	1	157441978	FY	Duffy blood-group antigen	1q	Duffy blood group antigen, absence	Homozygous*
CM942034	4	619702	PDE6B	Phosphodiesterase 6B, cGMP-specific, rod, beta	4p16.3	Retinitis pigmentosa 40	Heterozygous
CM021718	9	36208221	GNE	UDP-N-acetylglucosamine 2-epimerase	9p	Myopathy, distal, with rimmed vacuoles	Heterozygous
CM980633	10	50348375	ERCC6	Excision repair cross-complementing rodent repair deficiency, complementation group 6 protein (CSB)	10q	Cockayne syndrome	Homozygous†
CM050716	11	76531431	MYO7A	Myosin VIIA	11q13.5	Usher syndrome 1b	Homozygous†
CM950928	12	46812979	PFKM	Phosphofructokinase, muscle	12q13.3	Glycogen storage disease 7	Homozygous*
CM032029	14	20859880	RPGRIP1	Retinitis pigmentosa GTPase regulator interacting protein 1	14q11	Cone–rod dystrophy	Heterozygous
CM984025	19	18047618	IL12RB1	Interleukin-12 receptor, beta 1	19p13.1	Mycobacterial infection	Heterozygous
CM024138	19	41014441	NPHS1	Nephrosis-1, congenital, Finnish type	19q	Congenital nephrotic syndrome, Finnish type	Heterozygous
CM910052	22	49410905	ARSA	Arylsulphatase A	22q	Metachromatic leukodystrophy	Heterozygous

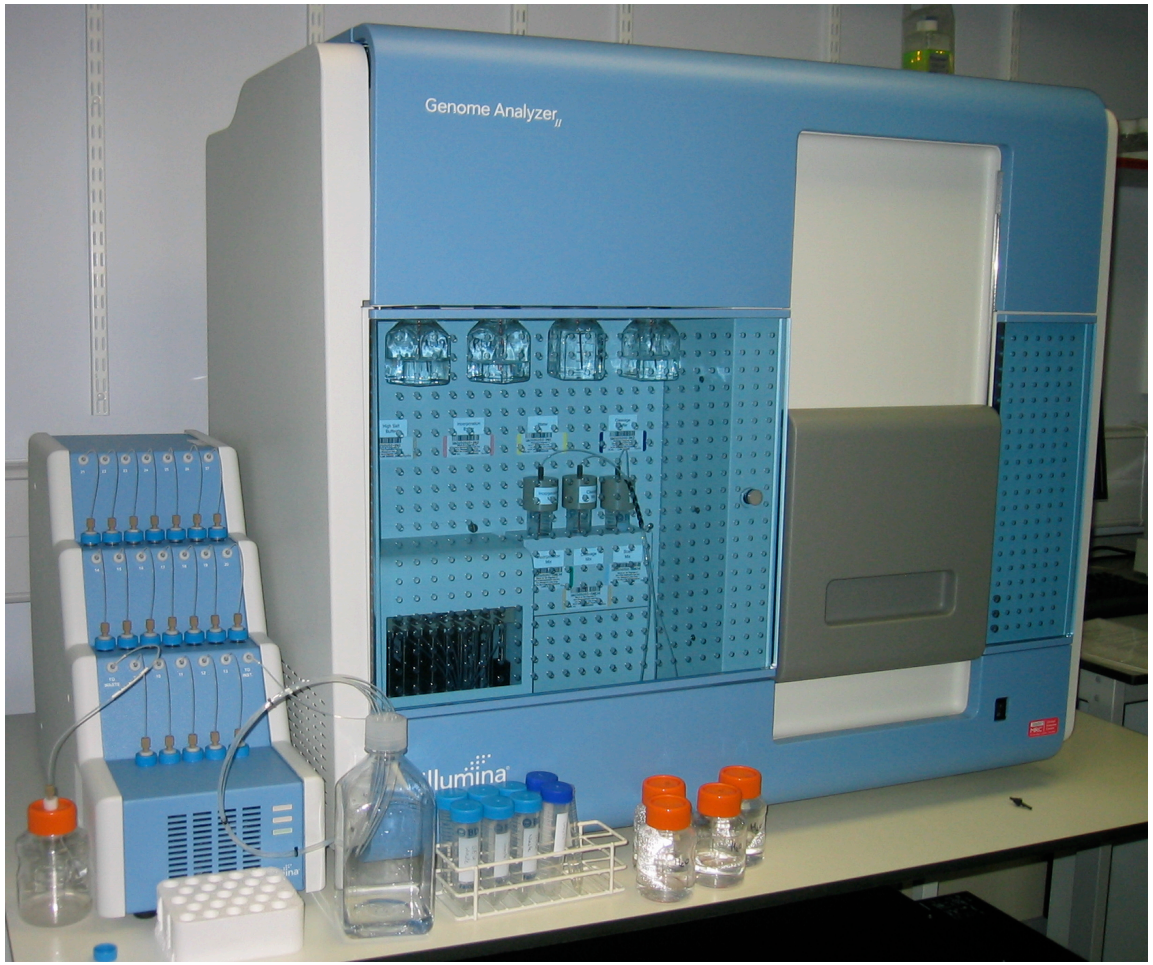
* Coverage at these SNP positions is less than 5. However, both produce benign phenotypes.
† Coverage at these SNP positions is greater than 5. Both would produce severe phenotypes if they were truly homozygous.

James Watson

Time:	~2 months
Coverage:	~7.4x
Sequence data:	106.5 million reads ~24.5 billion bases ~3.32 million SNPs ~260 runs on 454 instruments
Cost:	~$1 million

MRC Clinical Sciences Centre

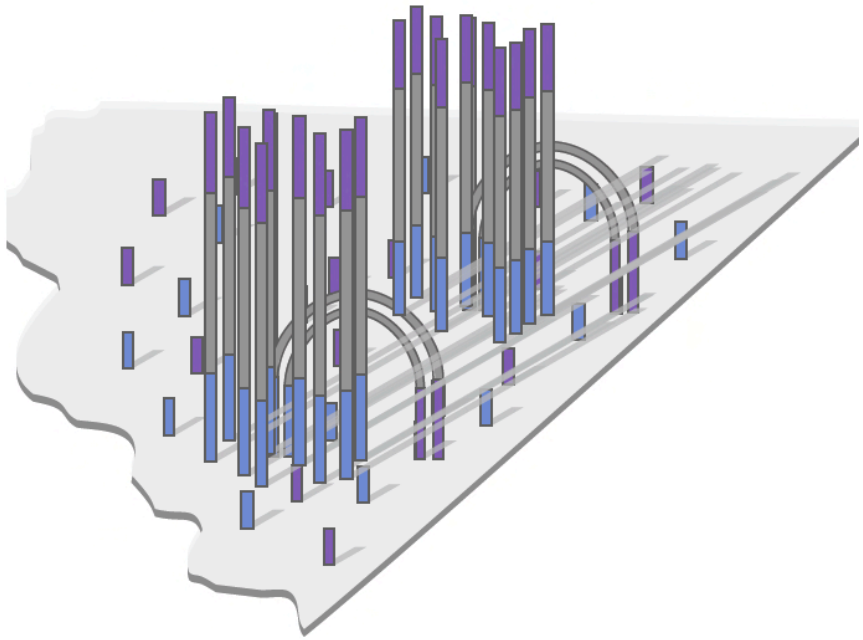
Solexa Genome Analyzer II



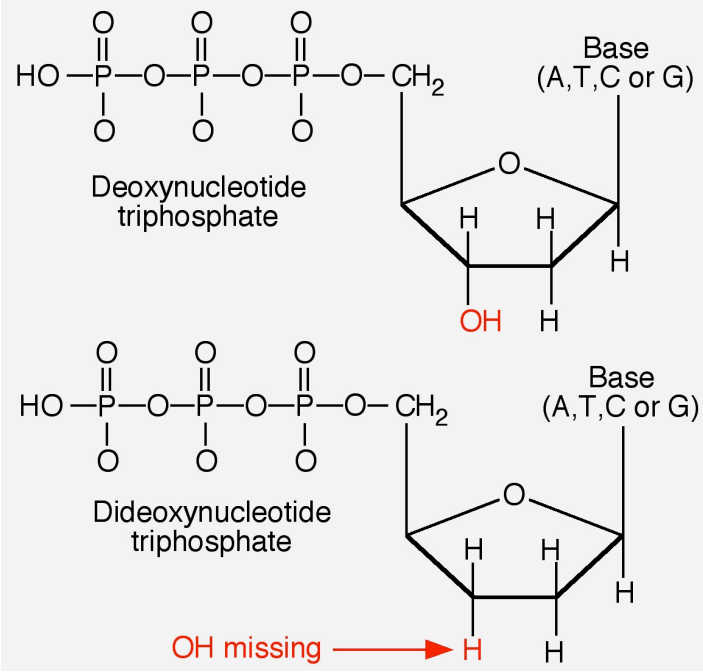
Solexa Genome Analyser II

Generate ssDNA fragments
~100 – 200 bp
Tagged with an ‘A’ and ‘B’ adapter

Amplified on Flow Cell by Bridge PCR to form clusters



Reversible Chain termination



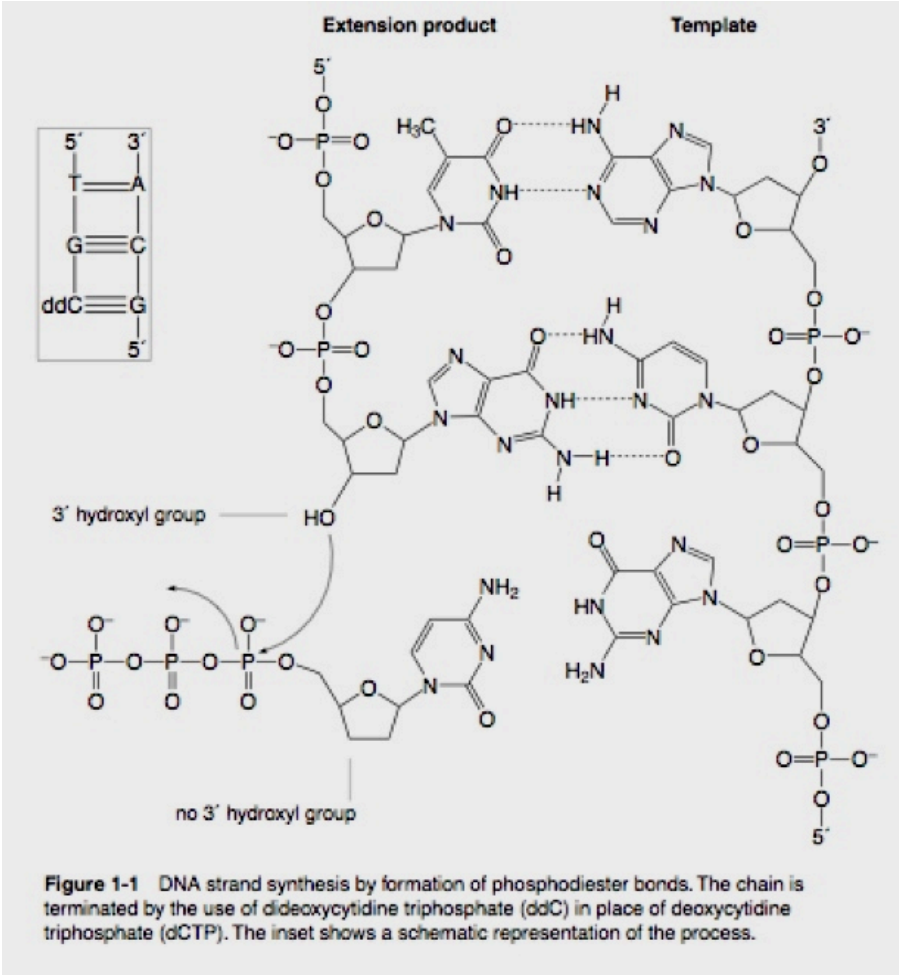
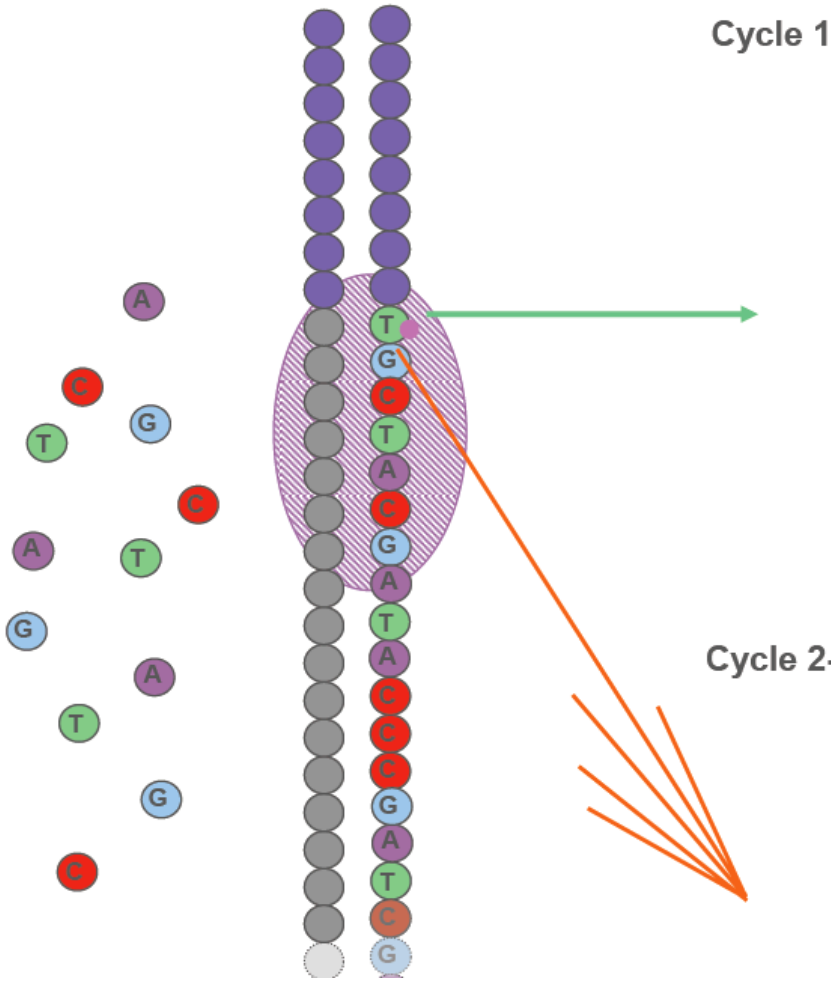


Figure 1-1 DNA strand synthesis by formation of phosphodiester bonds. The chain is terminated by the use of dideoxycytidine triphosphate (ddC) in place of deoxycytidine triphosphate (dCTP). The inset shows a schematic representation of the process.

Solexa Genome Analyser II



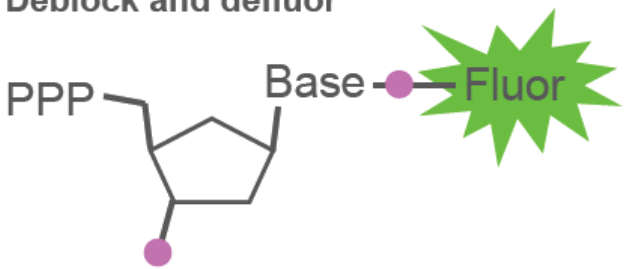
Cycle 1: Add sequencing reagents

First base incorporated

Remove unincorporated bases

Detect signal

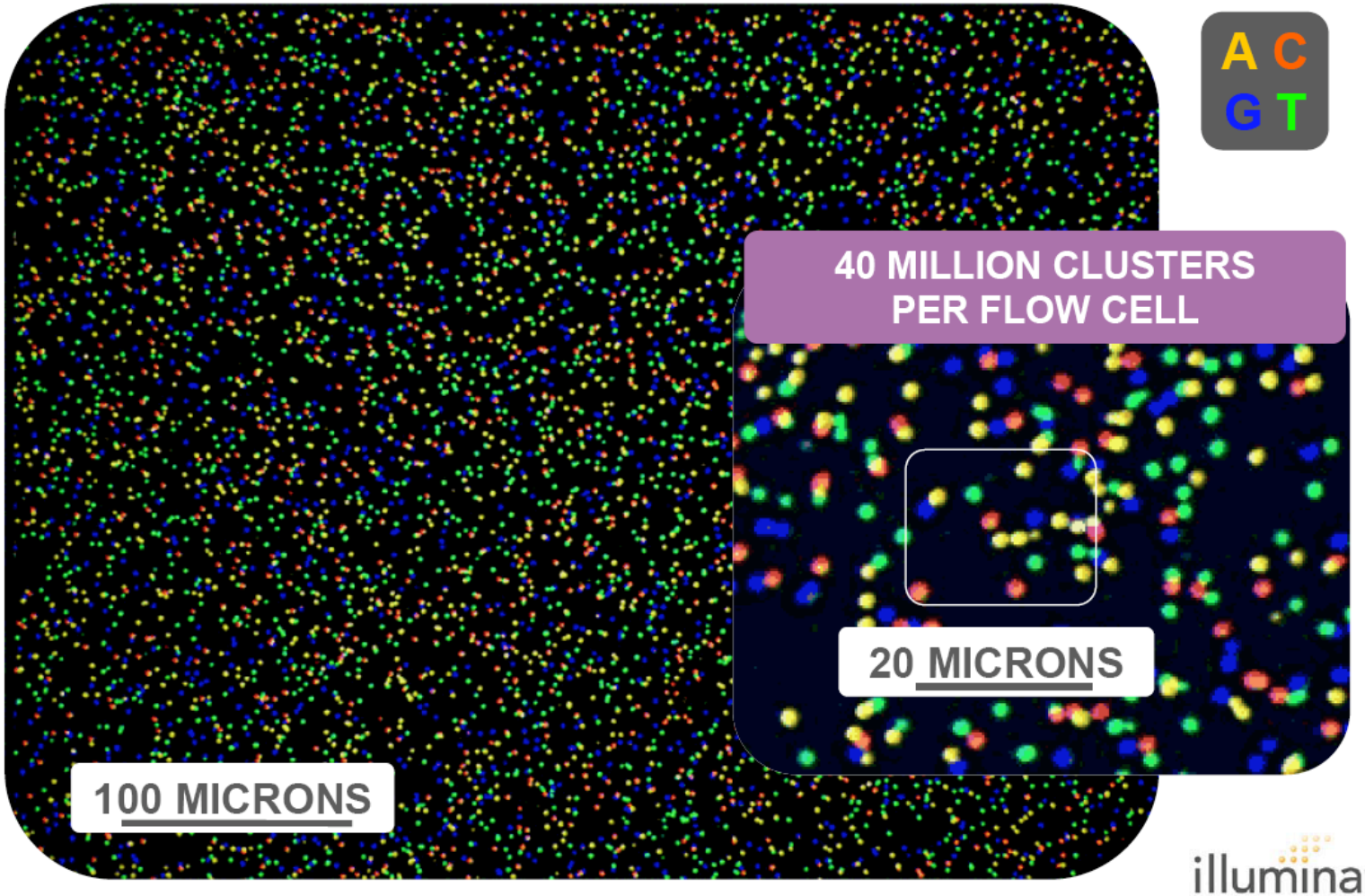
Deblock and defluor



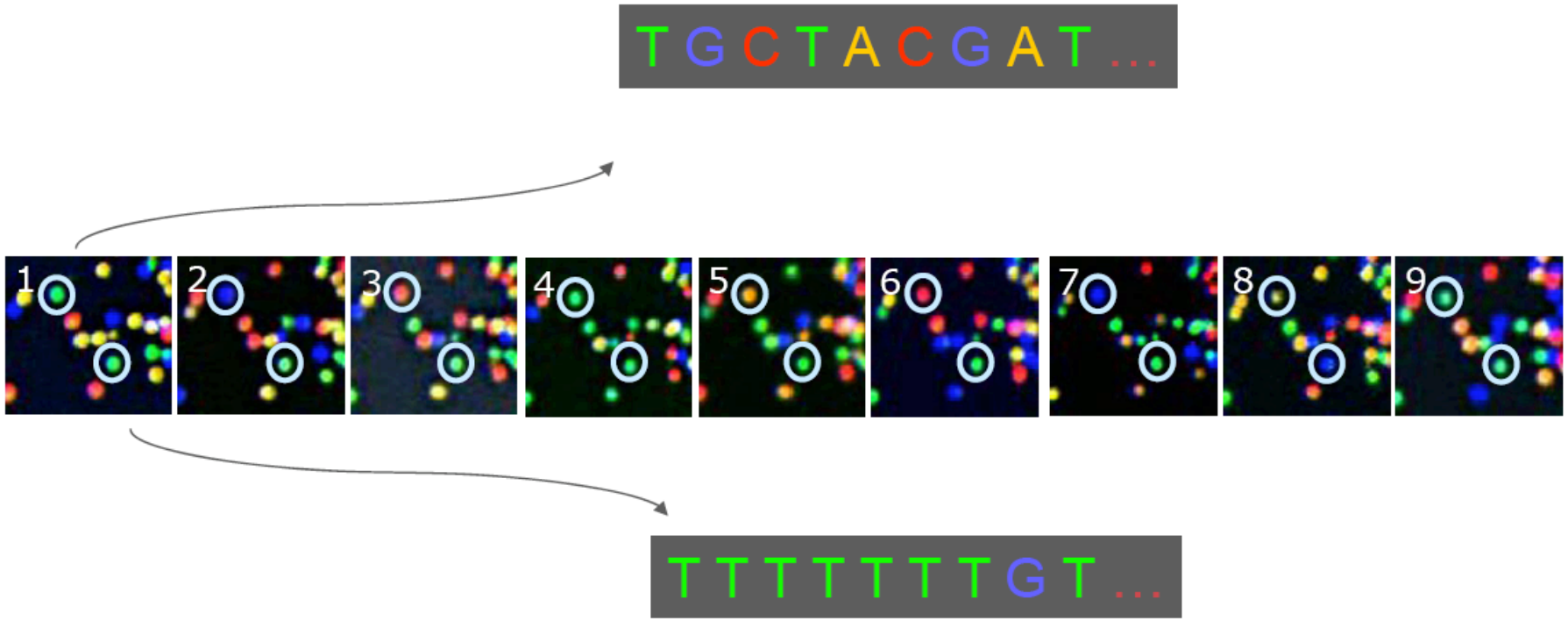
Cycle 2-n: Add sequencing reagents and repeat

- All four labelled nucleotides in one reaction
- Base-by-base sequencing
- No problems with homopolymer repeats
- High accuracy

Solexa Genome Analyser II



Solexa Genome Analyser II

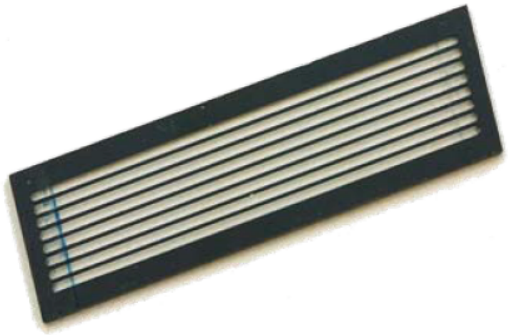


The New Technologies

Solexa HiSeq 2000

Read Length	*Run Time*	*Output*
2 x 50 bp	~ 5 days	~150 Gb
2 x 100 bp	~ 10 days	~300 Gb (~35 Gb/lane)

Reads: Up to 200 million clusters per lane



Tahiti and Fiji

Can sequence 8 complete human genomes in 2 weeks

MiSeq

2 x 150 base - 27 hours

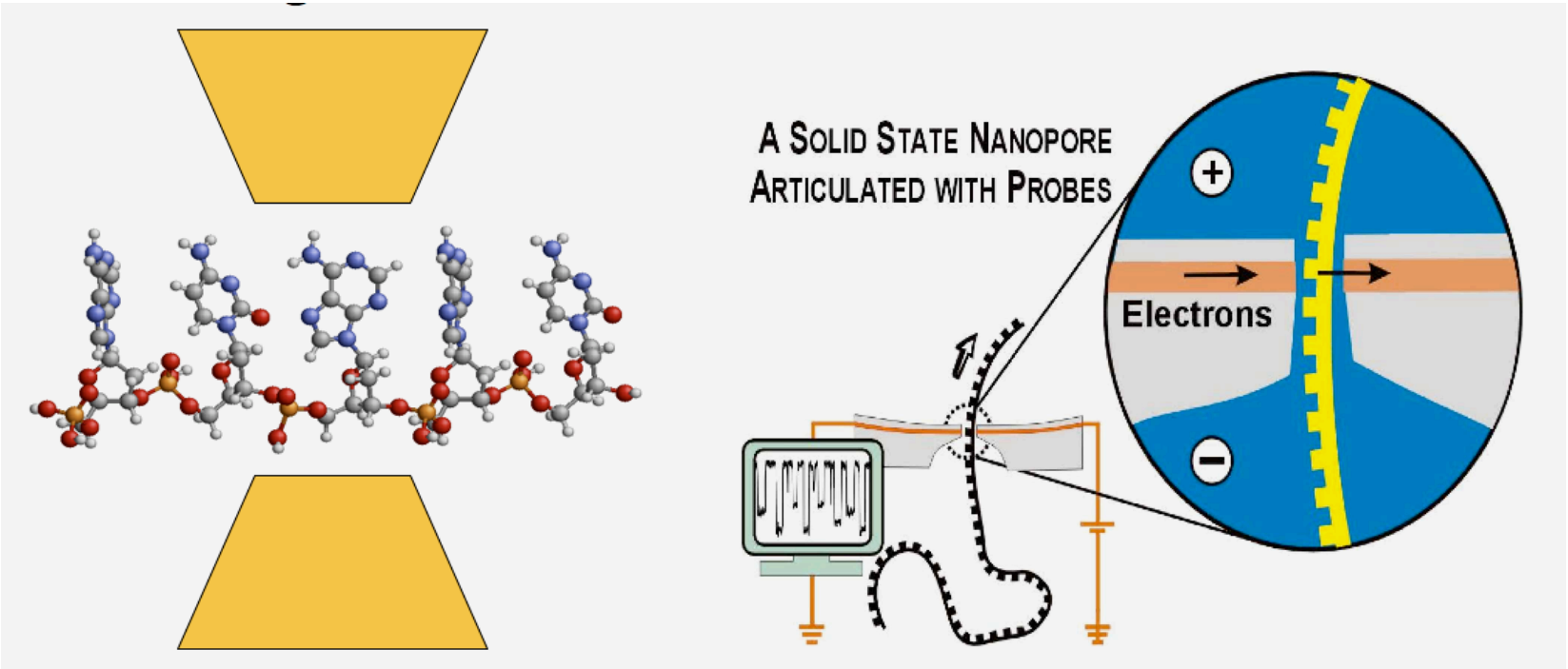
~1 gigabases

>3 million clusters

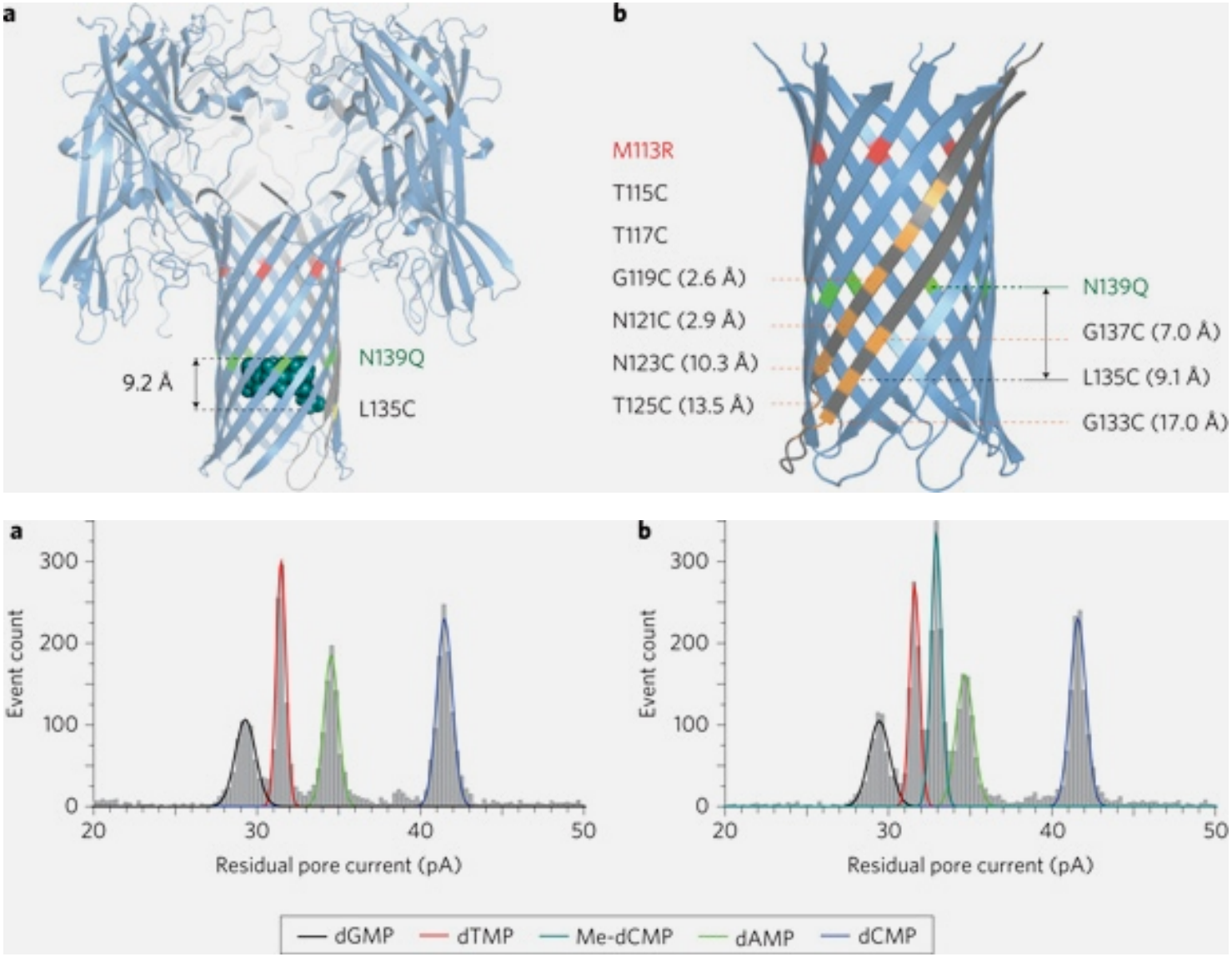
4 bacterial genomes to 50x coverage in 1 day

Single Molecule Nanopore Sequencing

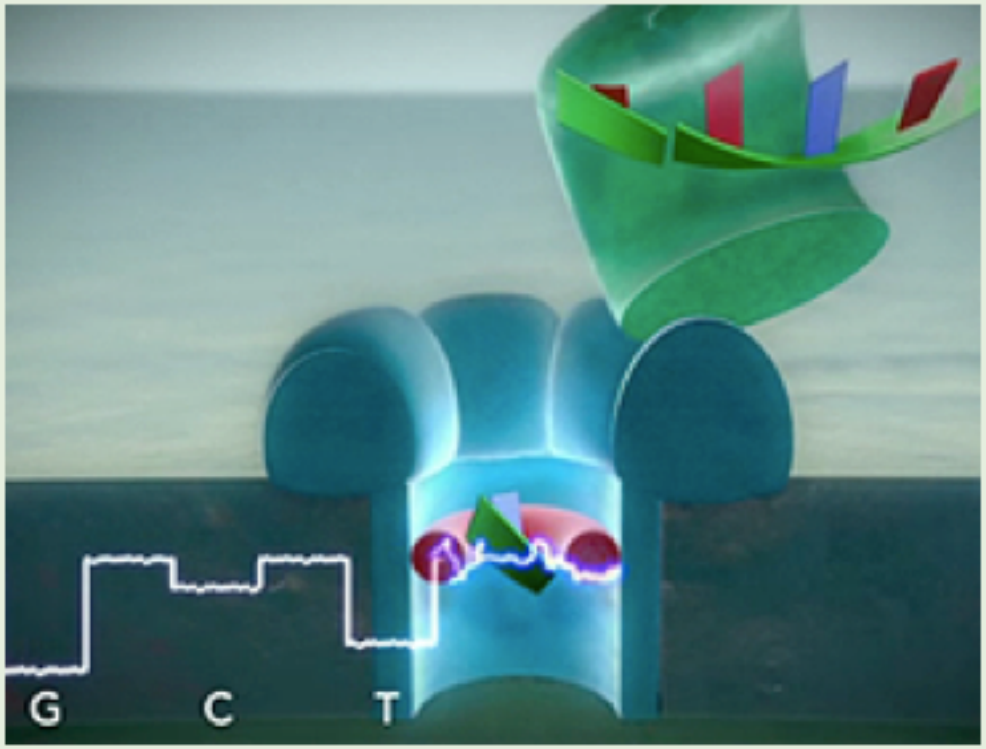
Preliminary Experimental Stage - 5 to 10 years before instruments available
However, incredible potential - complete genome in 1 experiment!



Oxford Nanopore Technologies Ltd

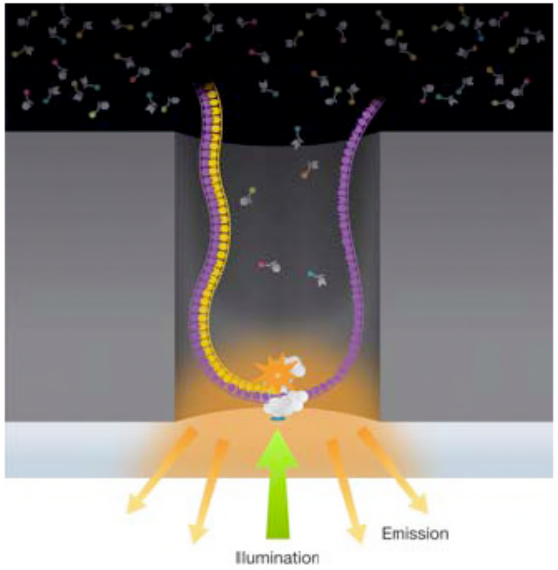


Oxford Nanopore Technologies Ltd

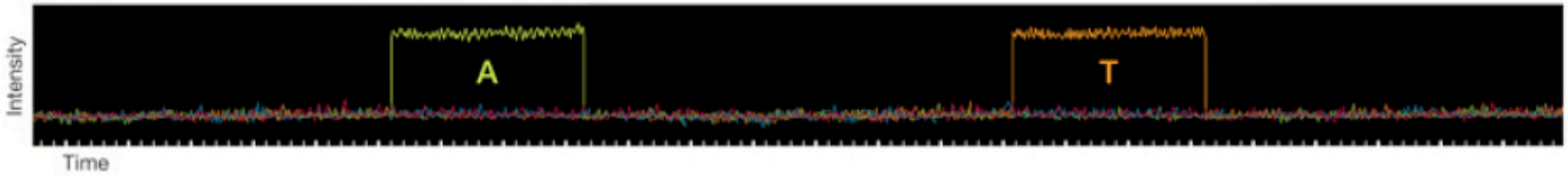


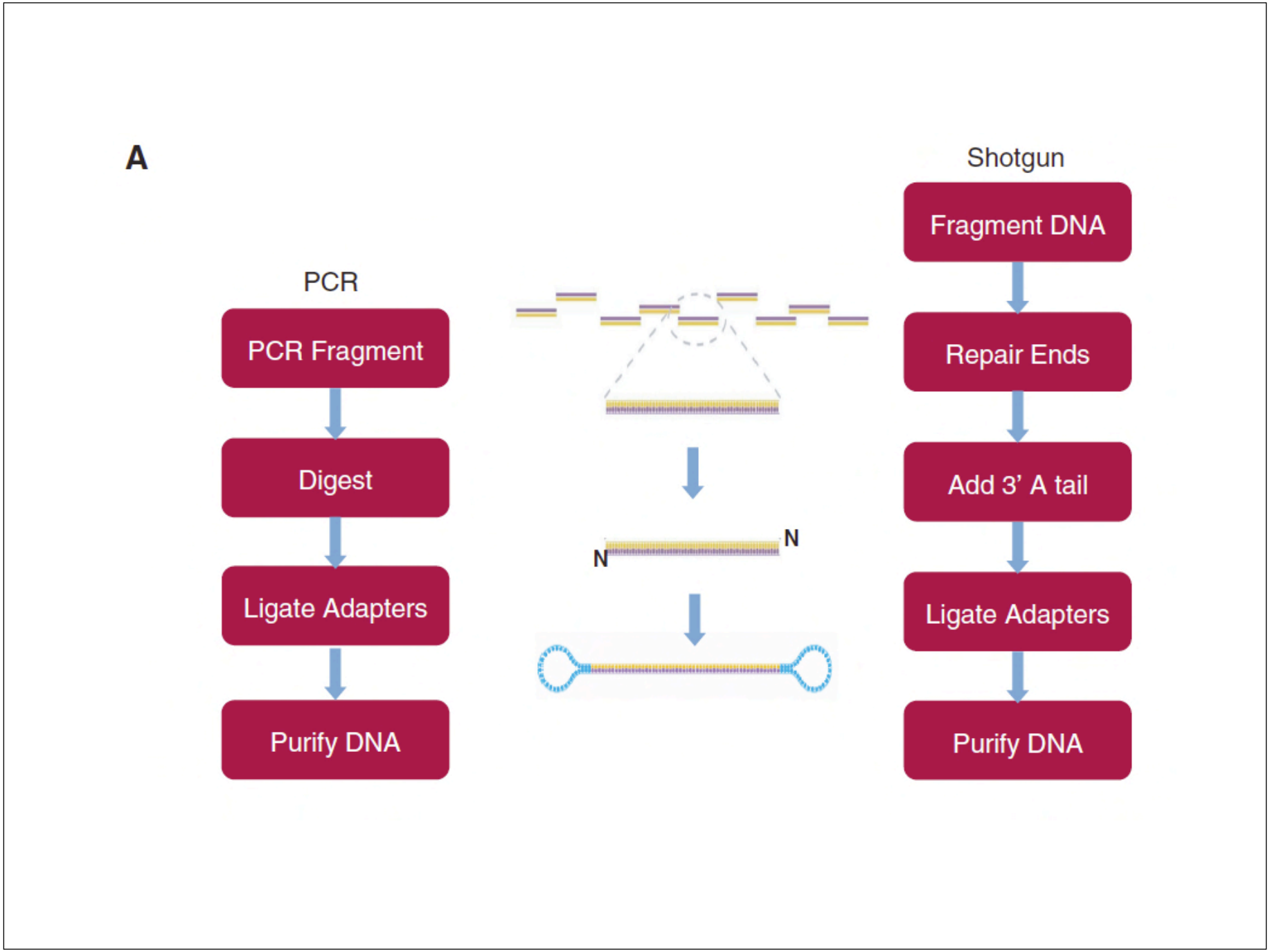
Nanopore

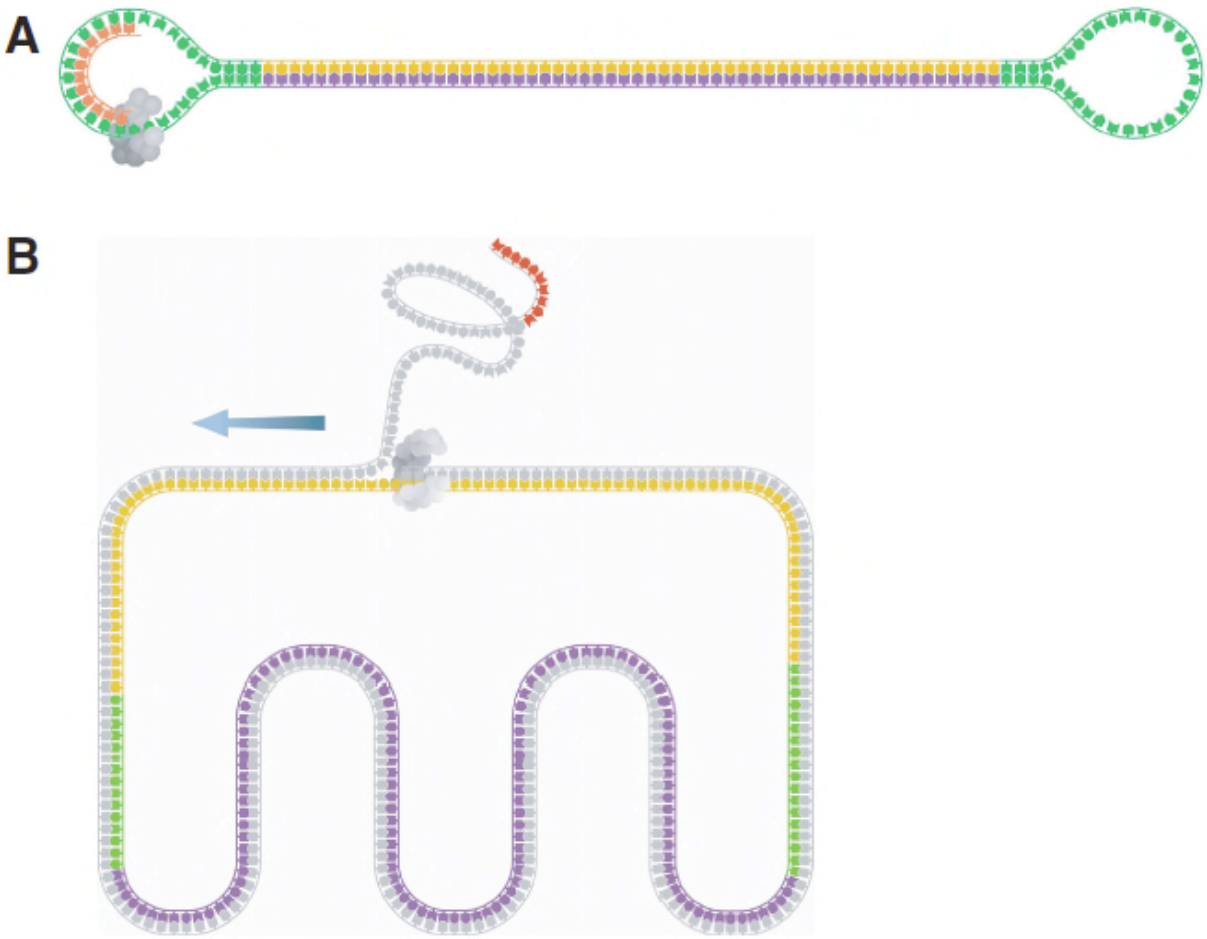
Real-Time Detection

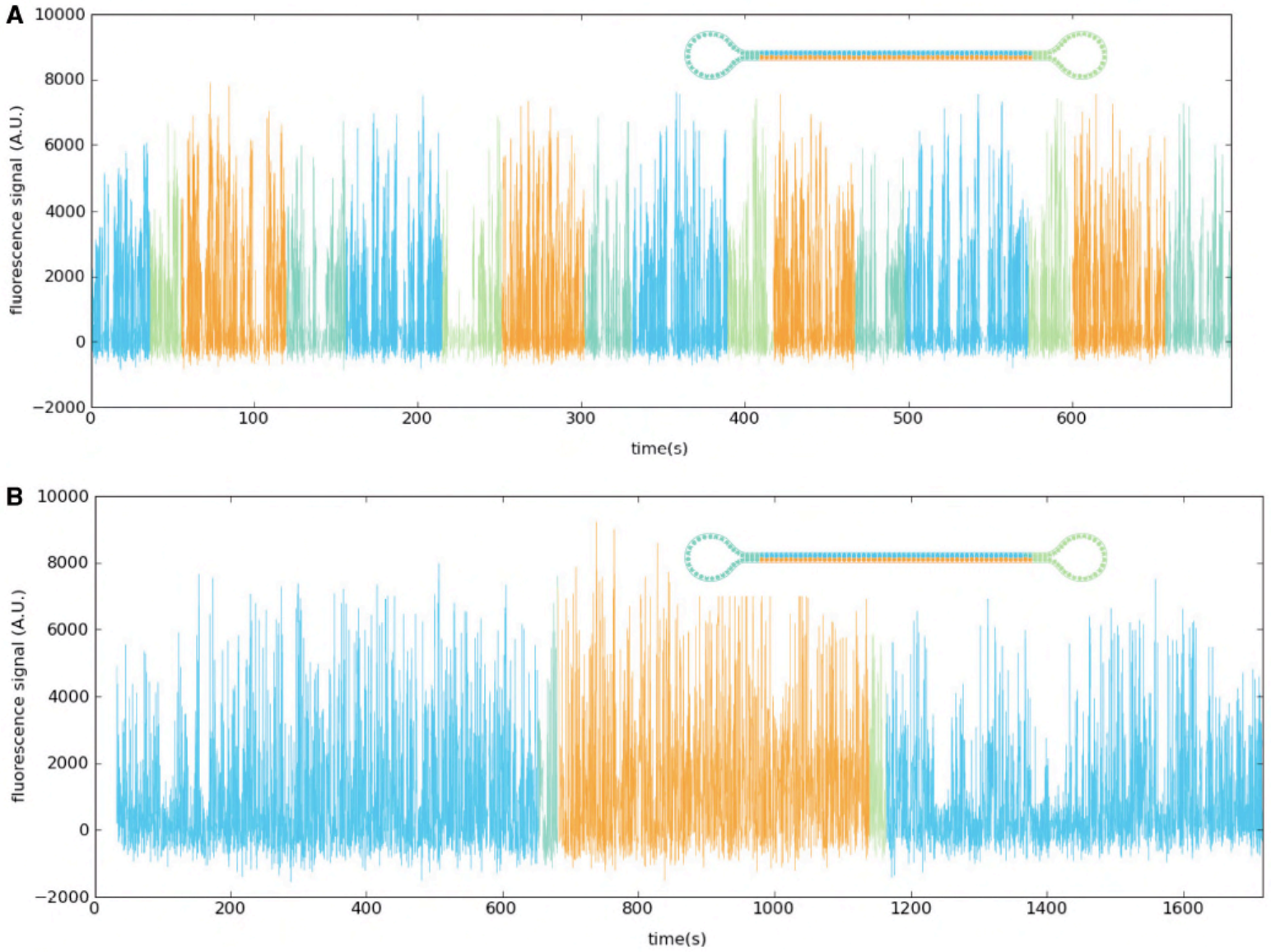


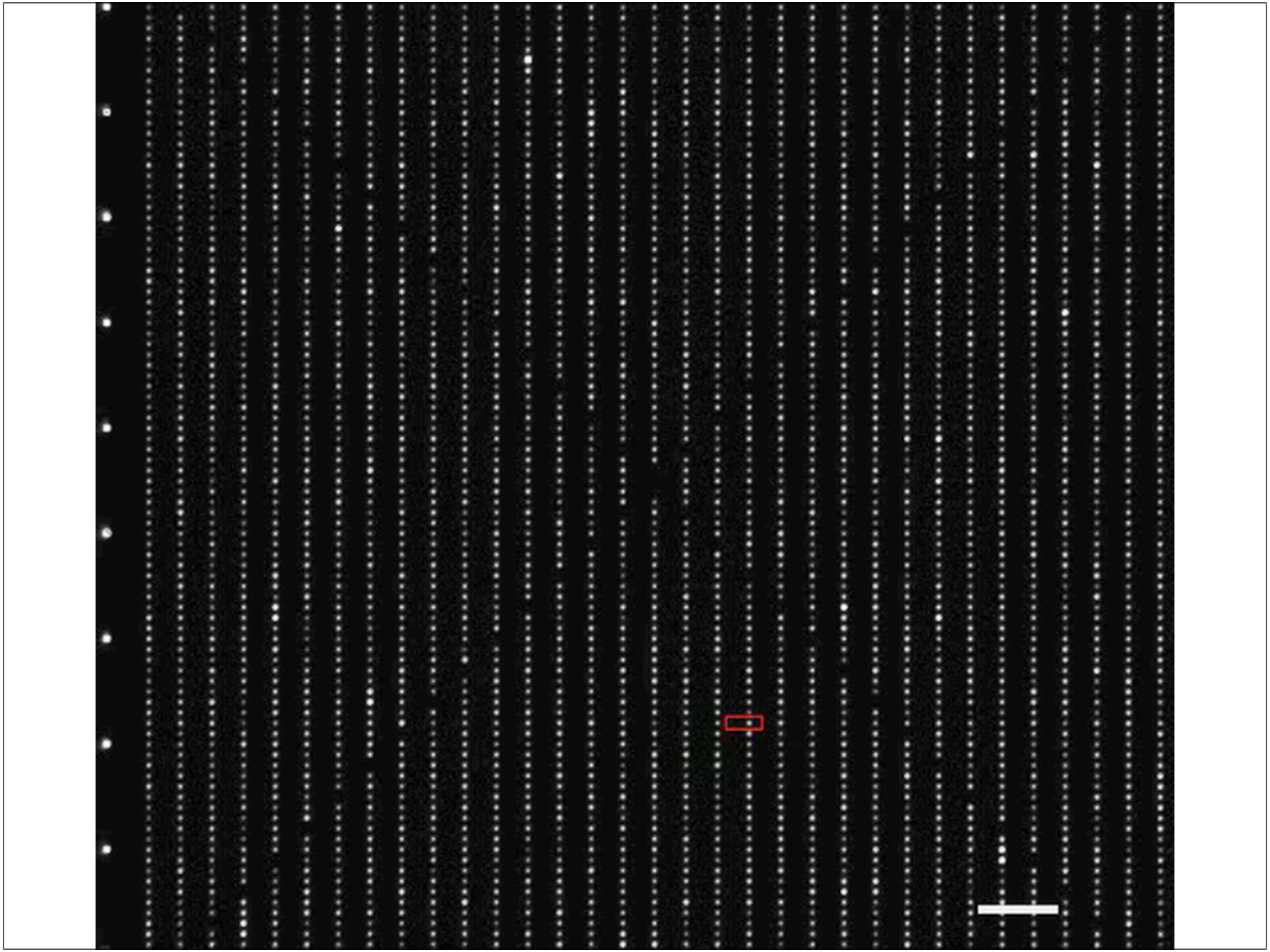
Zero-Mode Waveguides
~10 zeptolitres (10^{-21})




A
PCR
PCR Fragment
Digest
Ligate Adapters
Purify DNA
N
N
Shotgun
Fragment DNA
Repair Ends
Add 3' A tail
Ligate Adapters
Purify DNA


A
B


A
10000
8000
6000
4000
2000
0
−2000
fluorescence signal (A.U.)
0
100
200
300
400
500
600
time(s)
B
10000
8000
6000
4000
2000
0
−2000
fluorescence signal (A.U.)
0
200
400
600
800
1000
1200
1400
1600
time(s)



SMRT
Single Molecule Real Time

Prototype System:	80,000 ZMWs (Zero Mode Waveguides) 2 sets/SMRT Cell Read length ~1,000 bases/ZMW Speed ~2 bases/second

This yields: ~160 million bases/run in ~10 minutes

Version 2 System (2014):	1 million ZMWs/SMRT Read length up to 3,000 bases/ZMW Speed up to 50 bases/second

This yields: ~3 Gigabases in ~2 minutes

Length is irrelevant. Resolution depends upon the distance between each base, is the same at any point along a DNA chain.

Thus, could read a DNA molecule ~100,000 bases in size.

If had a chip with 1,000,000 pores, then can read ~100,000,000,000 bases (~30 fold coverage of the human genome).

Other Advantages:

1. No assembly issues - because of sequence read length

1. Most important - no cloning involved, so will get the regions not covered by Sanger Shotgun Method - possible methylated bases.

2. Potentially get complete chromosome sequences - No Gaps!

3. Sequence of both chromosomes - Mum & Dad

MRC Clinical Sciences Centre

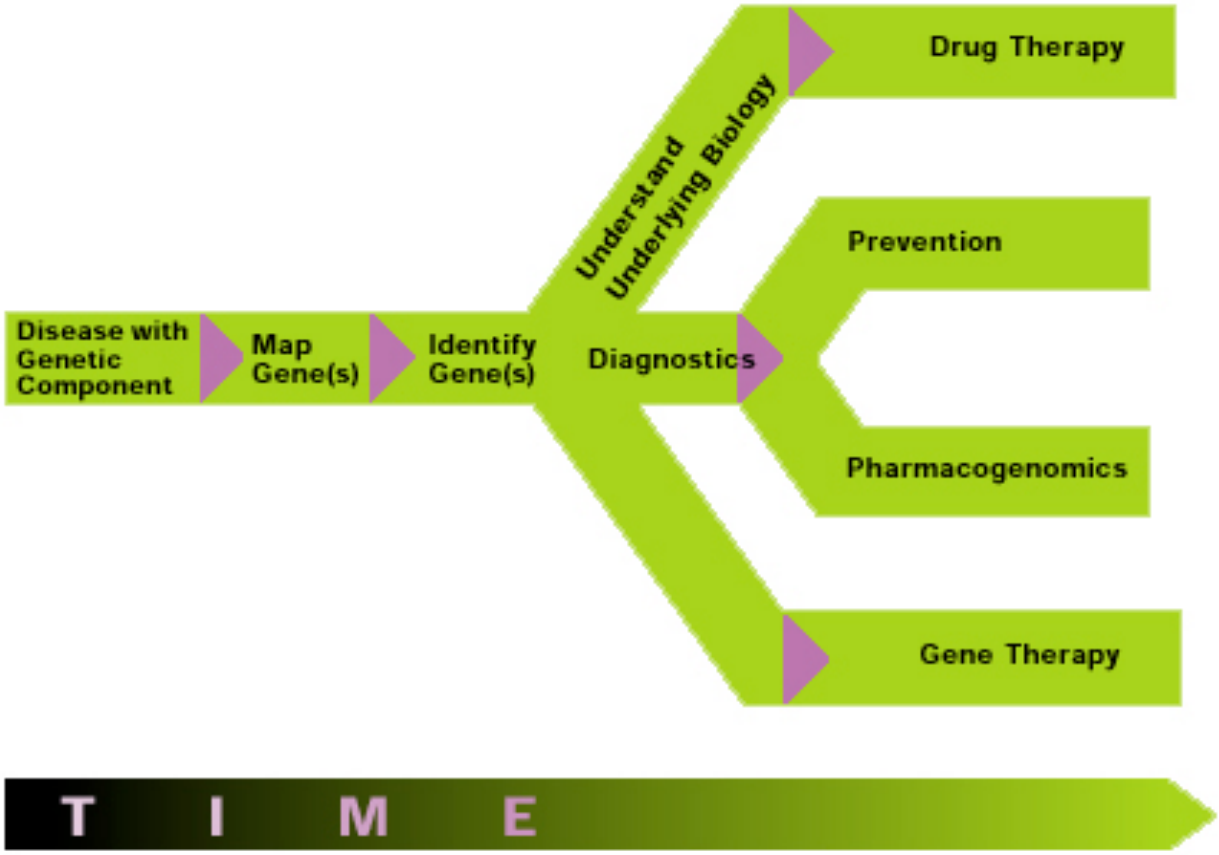
Timelines

Library Preparation:	Protocol dependent ~1 week
Roche – 454 Runs:	~24 hours (instrument time)
Solexa Runs:	~3 – 5 days (instrument time)
Data Analysis:	Eternity - *to infinity and beyond*

The Future

Pharmacogenomics Genetic or Genomic Medicine

Personalised medicines based on your genome



Pharmacogenomics

Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia.

McLeod HL, Krynetski EY, Relling MV, Evans WE.

Thiopurine methyltransferase (TPMT) catalyses the S-methylation of thiopurines, including 6-mercaptopurine and 6-thioguanine

TPMT activity exhibits genetic polymorphism, with about 1/300 inheriting TPMT deficiency as an autosomal recessive trait

Standard doses of thiopurines, TPMT-deficient patients accumulate excessive thioguanine nucleotides in hematopoietic tissues, leading to severe hematological toxicity that can be fatal

However, TPMT-deficient patients can be successfully treated with a 10- to 15-fold lower dosage of these medications